Parallel Gaussian Process Optimization with Upper Confidence Bound and Pure Exploration

Emile Contal, David Buffoni, Alexandre Robicquet, and Nicolas Vayatis

CMLA, ENS Cachan, CNRS, 61 Avenue du Président Wilson, F-94230 Cachan,
emile.contal@cmla.ens-cachan.fr

Abstract. In this paper, we consider the challenge of maximizing an unknown function f for which evaluations are noisy and are acquired with high cost. An iterative procedure uses the previous measures to actively select the next estimation of f which is predicted to be the most useful. We focus on the case where the function can be evaluated in parallel with batches of fixed size and analyze the benefit compared to the purely sequential procedure in terms of cumulative regret. We introduce the Gaussian Process Upper Confidence Bound and Pure Exploration algorithm (GP-UCB-PE) which combines the UCB strategy and Pure Exploration in the same batch of evaluations along the parallel iterations. We prove theoretical upper bounds on the regret with batches of size K for this procedure which show the improvement of the order of $\sqrt{K}$ for fixed iteration cost over purely sequential versions. Moreover, the multiplicative constants involved have the property of being dimension-free. We also confirm empirically the efficiency of GP-UCB-PE on real and synthetic problems compared to state-of-the-art competitors.

1 Introduction

Finding the maximum of a non-convex function by means of sequential noisy observations is a common task in numerous real world applications. The context of a high dimensional input space with expensive evaluation cost offers new challenges in order to come up with efficient and valid procedures. This problem of sequential global optimization arises for example in industrial system design and monitoring to choose the location of a sensor to find out the maximum response, or when determining the parameters of a heavy numerical code designed to maximize the output. The standard objective in this setting is to minimize the cumulative regret R_T, defined as the sum $\sum_{t=1}^{T} \left(f(x^\star) - f(x_t) \right)$ of the differences between the values of f at the points queried x_t and the true optimum of f noted $x^\star$. For a fixed horizon T, we refer to [20]. In the context where the horizon T is unknown, the query selection has to deal with the exploration/exploitation tradeoff. Successful algorithms have been developed in different settings to address this problem such as experimental design [15], Bayesian optimization [9, 19, 18, 31, 28, 27], active learning [8, 10], multiarmed bandit [2, 3, 11, 22, 24, 1, 33] and in particular Hierarchical Optimistic Optimization algorithm, HOO [6] for bandits in a generic space, namely $\mathcal{X}$-Armed bandits.

In some cases, it is possible to evaluate the function in parallel with batches of K queries with no increase in cost. This is typically the case in the sensors location problem if K sensors are available at each iteration, or in the numerical optimization problem on a cluster of K cores. Parallel strategies have been developed recently in [13, 4]. In the present paper, we propose to explore further the potential of parallel strategies for noisy function optimization with unknown horizon aiming simultaneously at practical efficiency and plausible theoretical results. We introduce a novel algorithm called GP-UCB-PE based on the Gaussian process approach which combines the benefits of the UCB policy with Pure Exploration queries in the same batch of K evaluations of f. The Pure Exploration component helps to reduce the uncertainty about f in order to support the UCB policy in finding the location of the maximum, and therefore in increasing the decay of the regret R_t at every timestep t. In comparison to other algorithms based on Gaussian processes and UCB such as GP-BUCB [13], the new algorithm discards the need for the initialization phase and offers a tighter control on the uncertainty parameter which monitors overconfidence. As a result, the derived regret bounds do not suffer from the curse of dimensionality since the multiplicative constants obtained are dimension free in contrast with the doubly exponential dependence observed in previous work. We also mention that Monte-Carlo simulations can be proposed as an alternative and this idea has been implemented in the *Simulation Matching* with UCB policy (SM-UCB) algorithm [4] which we also consider for comparison in the present paper. Unlike GP-BUCB, no theoretical guarantees for the SM-UCB algorithm are known for the bounds on the number of iterations needed to get close enough to the maximum, therefore the discussion will be reduced to empirical comparisons over several benchmark problems. The remainder of the paper is organized as follows. We state the background and our notations in Section 2. We formalize the Gaussian Process assumptions on f, and give the definition of regret in the parallel setting. We then describe the GP-UCB-PE algorithm and the main concepts in Section 3. We provide theoretical guarantees through upper bounds for the cumulative regret of GP-UCB-PE in Section 4. We finally show comparisons of our method and the related algorithms through a series of numerical experiments on real and synthetic functions in Section 5.

2   Problem Statement and Background

2.1   Sequential Batch Optimization

We address the problem of finding in the lowest possible number of iterations the maximum of an unknown function $f : \mathcal{X} \rightarrow \mathbb{R}$ where $\mathcal{X} \subset \mathbb{R}^d$, denoted by

$$f(x^{\star}) = \max_{x \in \mathcal{X}} f(x) \ .$$

The arbitrary choice of formulating the optimization problem as a maximization is without loss of generality, as we can obviously take the opposite of f if the problem is a minimization one. At each iteration t, we choose a batch of K points

in $\mathcal{X}$ called the queries $\{x_t^k\}_{0 \leqslant k < K}$, and then observe simultaneously the noisy values taken by f at these points,

$$y_t^k = f(x_t^k) + \epsilon_t^k \ ,$$

where the ϵ_t^k are independent Gaussian noise $\mathcal{N}(0, \sigma^2)$.

2.2   Objective

Assuming that the horizon T is unknown, a strategy has to be good at any iteration. We denote by $r_t^{(k)}$ the difference between the optimum of f and the point queried x_t^k,

$$r_t^{(k)} = f(x^\star) - f(x_t^k) \ .$$

We aim to minimize the cumulative regret,

$$R_T^K = \sum_{t < T} r_t^K \ ,$$

which is the standard objective with these formulations of the problem [5]. We focus on the case where the cost for a batch of evaluations of f is fixed. The loss r_t^K incurred at iteration t is then the simple regret for the batch [7], defined as

$$r_t^K = \min_{k < K} r_t^{(k)} \ .$$

An upper bound on R_T^K gives an upper bound of $\frac{R_T^K}{T}$ on the minimum gap between the best point found so far and the true maximum. We also provide bounds on the usual cumulative regret,

$$R_{TK} = \sum_{t < T} \sum_{k < K} r_t^{(k)} \ ,$$

which model the case where all the queries in a batch should have a low regret.

2.3   Gaussian Processes

In order to analyze the efficiency of a strategy, we have to make some assumptions on f. We want extreme variations of the function to have low probability.

Modeling f as a sample of a Gaussian Process (GP) is a natural way to formalize the intuition that nearby location are highly correlated. It can be seen as a continuous extension of multidimensional Gaussian distributions. We said that a random process f is Gaussian with mean function m and non-negative definite covariance function (kernel) k, written

$$f \sim GP(m, k) \ ,$$
$$\text{where } m : \mathcal{X} \to \mathbb{R}$$
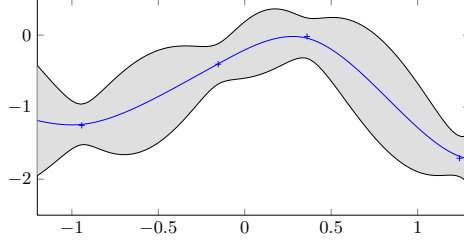$$\text{and } k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+ \ ,$$

Fig. 1. Gaussian Process inference of the posterior mean $\hat{\mu}$ (blue line) and deviation $\hat{\sigma}$ based on four realizations (blue crosses). The high confidence region (area in grey) is delimited by $\hat{f}^+$ and $\hat{f}^-$.

when for any finite subset of locations the values of the random function form a multivariate Gaussian random variable of mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ given by the mean m and the kernel k of the GP. That is, for all finite n and $x_1, \ldots, x_n \in \mathcal{X}$,

$$(f(x_1), \ldots, f(x_n)) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}) \ ,$$
$$\text{with } \boldsymbol{\mu}[x_i] = m(x_i)$$
$$\text{and } \mathbf{C}[x_i, x_j] = k(x_i, x_j) \ .$$

If we have the prior knowledge that f is drawn from a GP with zero mean [1] and known covariance (kernel) function, we can use Bayesian inference conditioned on the observations after T iterations to get the closed formulae for computing the posterior [30], which is a GP of mean and variance given at each location $x \in \mathcal{X}$ by,

$$\hat{\mu}_T(x) = \mathbf{k}_T(x)^\top \mathbf{C}_T^{-1} \mathbf{Y}_T \tag{1}$$
$$\text{and } \hat{\sigma}_T^2(x) = k(x, x) - \mathbf{k}_T(x)^\top \mathbf{C}_T^{-1} \mathbf{k}_T(x) \ , \tag{2}$$

$\mathbf{X}_T = \{x_t^k\}_{t<T, k<K}$ is the set of queried locations, $\mathbf{Y}_T = [y_t^k]_{x_t^k \in \mathbf{X}_T}$ is the vector of noisy observations, $\mathbf{k}_T(x) = [k(x_t^k, x)]_{x_t^k \in \mathbf{X}_T}$ is the vector of covariances between x and the queried points, and $\mathbf{C}_T = \mathbf{K}_T + \sigma^2 \mathbf{I}$ with $\mathbf{K}_T = [k(x, x')]_{x, x' \in \mathbf{X}_T}$ the kernel matrix and $\mathbf{I}$ stands for the identity matrix.

The three most common kernel functions are:

- the polynomial kernels of degree $\alpha \in \mathbb{N}$, $k(x_1, x_2) = (x_1^\top x_2 + c)^\alpha$, $c \in \mathbb{R}$
- the (Gaussian) Radial Basis Function kernel (RBF or Squared Exponential) with length-scale $l > 0$, $k(x_1, x_2) = \exp\left(-\frac{\|x_1, x_2\|^2}{2l^2}\right)$,
- the Matérn kernel, of length-scale l and parameter ν,

$$k(x_1, x_2) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\, \|x_1, x_2\|}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\, \|x_1, x_2\|}{l}\right) \ , \tag{3}$$

where K_ν is the modified Bessel function of the second kind and order ν.

---

[1] this is without loss of generality as the kernel k can completely define the GP [30].

The Bayesian inference is represented on Figure 1 in a sample problem in dimension 1. The posteriors are based on four observations of a Gaussian Process. The vertical height of the grey area is proportional to the posterior deviation at each point.

3 Parallel Optimization Procedure

3.1 Confidence Region

A key property from the GP framework is that the posterior distribution at a location x has a normal distribution $\mathcal{N}(\widehat{\mu}_T(x), \widehat{\sigma}_T^2(x))$. We can then define a upper confidence bound $\widehat{f}^+$ and a lower confidence bound $\widehat{f}^-$, such that f is included in the interval with high probability,

$$\widehat{f}_T^+(x) = \widehat{\mu}_T(x) + \sqrt{\beta_T}\widehat{\sigma}_{T-1}(x) \tag{4}$$

$$\text{and } \widehat{f}_T^-(x) = \widehat{\mu}_T(x) - \sqrt{\beta_T}\widehat{\sigma}_{T-1}(x) \ , \tag{5}$$

with $\beta_T \in \mathcal{O}(\log T)$ defined in Section 4.

$\widehat{f}^+$ and $\widehat{f}^-$ are illustrated on Figure 1 respectively by the upper and lower envelope of the grey area. The region delimited in that way, the high confidence region, contains the unknown f with high probability. This statement will be a main element in the theoretical analysis of the algorithm in Section 4.

3.2 Relevant Region

We define the relevant region $\mathfrak{R}_t$ being the region which contains $x^\star$ with high probability. Let $y_t^\bullet$ be our lower confidence bound on the maximum,

$$y_t^\bullet = \widehat{f}_t^-(x_t^\bullet), \text{ where } x_t^\bullet = \operatorname*{argmax}_{x \in \mathcal{X}} \widehat{f}_t^-(x) \ .$$

$y_t^\bullet$ is represented by the horizontal dotted green line on Figure 2. $\mathfrak{R}_t$ is defined as,

$$\mathfrak{R}_t = \left\{ x \in \mathcal{X} \mid \widehat{f}_t^+(x) \geqslant y_t^\bullet \right\} \ .$$

$\mathfrak{R}_t$ discard the locations where $x^\star$ does not belong with high probability. It is represented in green on Figure 2. We refer to [17] for related work in the special case of deterministic Gaussian Process Bandits.

In the sequel, we will use a modified version of the relevant region which also contains $\operatorname{argmax}_{x \in \mathcal{X}} \widehat{f}_{t+1}^+(x)$ with high probability. The novel relevant region is formally defined by

$$\mathfrak{R}_t^+ = \left\{ x \in \mathcal{X} \mid \widehat{\mu}_t(x) + 2\sqrt{\beta_{t+1}}\widehat{\sigma}_{t-1}(x) \geqslant y_t^\bullet \right\} \ . \tag{6}$$

Using $\mathfrak{R}_t^+$ instead of $\mathfrak{R}_t$ guarantees that the queries at iteration t will leave an impact on the future choices at iteration $t + 1$.

Algorithm 1: GP-UCB-PE

---

   for $t = 0, \ldots, T$ **do**
        Compute $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ with Eq.1 and Eq.2
        $x_t^0 \leftarrow \mathrm{argmax}_{x \in \mathcal{X}} \, \widehat{f}_t^+(x)$
        Compute $\mathfrak{R}_t^+$ with Eq.6
        for $k = 1, \ldots, K - 1$ **do**
               Compute $\widehat{\sigma}_t^{(k)}$ with Eq.2
               $x_t^k \leftarrow \mathrm{argmax}_{x \in \mathfrak{R}_t^+} \, \widehat{\sigma}_t^{(k)}(x)$
        Query $\{x_t^k\}_{k < K}$

3.3 GP-UCB-PE

We present here the Gaussian Process Upper Confidence Bound with Pure Exploration algorithm, GP-UCB-PE, a novel algorithm combining two strategies to determine the queries $\{x_t^k\}_{k < K}$ for batches of size K. The first location is chosen according to the GP-UCB rule,

$$x_t^0 = \underset{x \in \mathcal{X}}{\mathrm{argmax}} \, \widehat{f}_t^+(x) \; . \tag{7}$$

This single rule is enough to tackle the exploration/exploitation tradeoff. The value of β_t balances between exploring uncertain regions (high posterior variance $\widehat{\sigma}_t^2(x)$) and focusing on the supposed location of the maximum (high posterior mean $\widehat{\mu}_t(x)$). This policy is illustrated with the point x^0 on Figure 2.

The $K - 1$ remaining locations are selected via Pure Exploration restricted to the region $\mathfrak{R}_t^+$. We aim to maximize $I_t(\mathbf{X}_t^{K-1})$, the information gain about f by the locations $\mathbf{X}_t^{K-1} = \{x_t^k\}_{1 \leqslant k < K}$ [12]. Formally, $I_t(\mathbf{X})$ is the gain in entropy when knowing the values of the observations $\mathbf{Y}$ at $\mathbf{X}$, conditioned on $\mathbf{X}_t$ the observations we have seen so far.

$$I_t(\mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}_t) \; . \tag{8}$$

Finding the $K - 1$ points that maximize I_t for any integer K is known to be NP-complete [23]. However, due to the submodularity of I_t [19], it can be efficiently approximated by the greedy procedure which selects the points one by one and never backtracks. For a Gaussian distribution, $H(\mathcal{N}(\mu, \mathbf{C})) = \frac{1}{2} \log \det(2\pi e \mathbf{C})$. We thus have $I_t(\mathbf{X}) \in \mathcal{O}(\log \det \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{X}$. For GP, the location of the single point that maximizes the information gain is easily computed by maximizing the posterior variance. For all $1 \leqslant k < K$ our greedy strategy selects the following points one by one,

$$x_t^k = \underset{x \in \mathfrak{R}_t^+}{\mathrm{argmax}} \, \widehat{\sigma}_t^{(k)}(x) \; , \tag{9}$$

where $\widehat{\sigma}_t^{(k)}$ is the updated variance after choosing $\{x_t^{k'}\}_{k' < k}$. We use here the fact that the posterior variance does not depend on the values y_t^k of the observations,
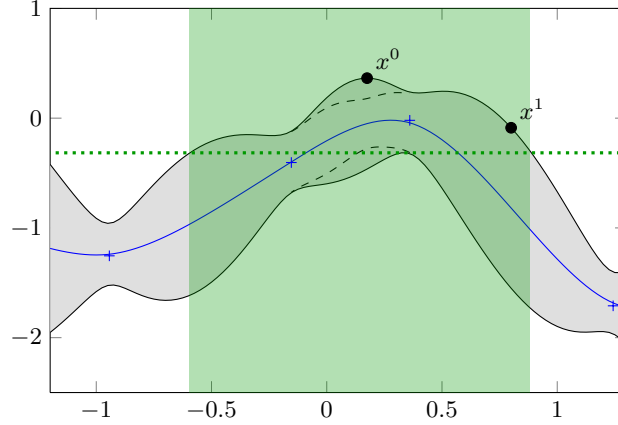
Fig. 2. Two queries of GP-UCB-PE on the previous example. The lower confidence bound on the maximum is represented by the horizontal dotted green line at $y_t^\bullet$. The relevant region $\mathfrak{R}$ is shown in light green (without edges). The first query x^0 is the maximizer of $\widehat{f}^+$. We show in dashed line the upper and lower bounds with the update of $\widehat{\sigma}$ after having selected x^0. The second query x^1 is the one maximizing the uncertainty inside $\mathfrak{R}^+$, an extension of $\mathfrak{R}$ which is not illustrated here.

but only on their position x_t^k. One such point is illustrated with x^1 on Figure 2. These $K - 1$ locations reduce the uncertainty about f, improving the guesses of the UCB procedure by x_t^0. The overall procedure is shown in Algorithm 1.

3.4   Numerical Complexity

Even if the numerical cost of GP-UCB-PE is insignificant in practice compared to the cost of the evaluation of f, the complexity of the exact computations of the variances (Eq.2) is in $\mathcal{O}(n^3)$ and might by prohibitive for large $n = tK$. One can reduce drastically the computation time by means of Lazy Variance Calculation [13], built on the fact that $\widehat{\sigma}_t(x)$ always decreases when t increases for all $x \in \mathcal{X}$. We further mention that efficient approximated inference algorithms such as the EP approximation and MCMC sampling [26] can be used in order to face the challenge of large n.

4   Regret Bounds

4.1   Main Result

The main theoretical result of this article is the upper bound on the regret formulated in Theorem 1. We need to adjust the parameter β_t such that $f(x)$ is contained by the high confidence region for all iterations t with probability at least $1 - \delta$ for a fixed $0 < \delta < 1$.

- If $\mathcal{X}$ is finite, then we choose $\beta_t = 2\log(|\mathcal{X}|\frac{\pi_t}{\delta})$ where $\pi_t > 0$ such that $\sum_{t=0}^{\infty} \pi_t^{-1} = 1$. We set for example $\beta_t = 2\log\left(|\mathcal{X}|\, t^2 \frac{\pi^2}{6\delta}\right)$.
- If $\mathcal{X} \subset [0,r]^d$ is compact and convex, we need the following bounds on the derivatives of f,

$$\exists a,b > 0, \ \forall j \leqslant d, \ \ \Pr\left(\sup_{x\in\mathcal{X}}\left|\frac{\partial f}{\partial x_j}\right| > L\right) \leqslant ae^{-\frac{L^2}{b^2}} \ .$$

Then, we can set

$$\beta_t = 2\log\left(t^2\frac{2\pi^2}{3\delta}\right) + 2d\log\left(t^2 dbr\sqrt{\log\left(\frac{4da}{\delta}\right)}\right) \ .$$

The regret bound are expressed in term of γ_{TK}, the maximum information gain (Eq. 8) obtainable by a sequence of TK queries.

$$\gamma_t = \max_{\mathbf{X}\subset\mathcal{X}, |\mathbf{X}|=t} I_0(\mathbf{X}) \ .$$

Under these assumptions, we obtain the following result.

Theorem 1. *Fix $0 < \delta < 1$ and consider the calibration of β_t defined as above, assuming $f \sim GP(0,k)$ with bounded variance, $\forall x \in \mathcal{X}, \ k(x,x) \leqslant 1$, then the cumulative regret incurred by* **GP-UCB-PE** *on f is bounded by $\mathcal{O}\left(\sqrt{\frac{T}{K}\beta_T\gamma_{TK}}\right)$ whp, More precisely, with $C_1 = \frac{4}{\log(1+\sigma^{-2})}$, and $C_2 = \frac{\pi}{\sqrt{6}}$, $\forall T$,*

$$\Pr\left(R_T^K \leqslant \sqrt{C_1\frac{T}{K}\beta_T\gamma_{TK} + C_2}\right) \geqslant 1 - \delta \ .$$

For R_{TK} we obtain similar bounds with $C_1 = \frac{36}{\log(1+\sigma^{-2})}$,

$$\Pr\left(R_{TK} \leqslant \sqrt{C_1 T\beta_T\gamma_{TK} + C_2}\right) \geqslant 1 - \delta \ .$$

4.2   Discussion

When $K \ll T$, the probabilistic bound is better than the one of sequential GP-UCB by an order of $\sqrt{K}$. Compared to [13], we remove the need of the initialization phase. GP-UCB-PE does not need either to multiply the uncertainty parameter β_t by $\exp(\gamma_{TK}^{\text{init}})$ where $\gamma_{TK}^{\text{init}}$ is equal to the maximum information gain obtainable by a sequence of TK queries after the initialization phase. The improvement can be doubly exponential in the dimension d in the case of RBF Kernels. To the best of our knowledge, no regret bounds have been proven for the *Simulation Matching* algorithm.

The values of γ_{TK} for different common kernel are reported in Table 1, where d is the dimension of the space considered and $\alpha = \frac{d(d+1)}{2\nu+d(d+1)} \leqslant 1$, ν being the Matérn parameter. We also compare on Table 1 the general forms of the bounds for the regret obtained by GP-UCB-PE and GP-BUCB up to constant terms. The cumulative regret we obtained with RBF Kernel is of the form $\tilde{\mathcal{O}}\left(\sqrt{\frac{T}{K}(\log TK)^d}\right)$ against $\tilde{\mathcal{O}}\left(\exp((\frac{2d}{e})^d)\sqrt{\frac{T}{K}(\log TK)^d}\right)$ for GP-BUCB.

Table 1. General Forms of Regret Bounds for GP-UCB-PE and GP-BUCB

	GP-UCB-PE	GP-BUCB
R_T^K	$\sqrt{\frac{T \log T}{K}} \gamma_{TK}$	$C\sqrt{\frac{T \log TK}{K}} \gamma_{TK}$

Kernel	Linear	RBF	Matérn
γ_{TK}	$d \log TK$	$(\log TK)^{d+1}$	$(TK)^{\alpha} \log TK$
C	$\exp(\frac{2}{e})$	$\exp((\frac{2d}{e})^d)$	e

4.3   Proofs of the main result

In this section, we analyze theoretically the regret bounds for the GP-UCB-PE algorithm. We provide here the main steps for the proof of Theorem 1. On one side the UCB rule of the algorithm provides a regret bounded by the information we have on f conditioned on the values observed so far. On the other side, the Pure Exploration part gathers information and therefore accelerates the decrease in uncertainty. We refer to [13] for the proofs of the bounds for GP-BUCB.

For the sake of concision, we introduce the notations σ_t^k for $\hat{\sigma}_t^{(k)}(x_t^k)$ and σ_t^0 for $\hat{\sigma}_{t-1}(x_t^0)$. We simply bound r_t^K the regret for the batch at iteration t by the simple regret $r_t^{(0)}$ for the single query chosen via the UCB rule. We then give a bound for $r_t^{(0)}$ which is proportional to the posterior deviations σ_t^0. Knowing that the sum of all $(\sigma_t^k)^2$ is not greater than $C_1 \gamma_{TK}$, we want to prove that the sum of the $(\sigma_t^0)^2$ is less than this bound divided by K. The arguments are based on the fact that the posterior for $f(x)$ is Gaussian, allowing us to choose β_t such that

$$\forall x \in \mathcal{X}, \forall t < T, \ f(x) \in [\hat{f}_t^-(x), \hat{f}_t^+(x)]$$

holds with high probability. Here and in the following, "with high probability" or *whp* means "with probability at least $1 - \delta$" for any $0 < \delta < 1$, the definition of β_t being dependent of δ.

Lemma 1. *For finite $\mathcal{X}$, we have $r_t^K \leqslant r_t^{(0)} \leqslant 2\sqrt{\beta_t}\sigma_t^0$, and for compact and convex $\mathcal{X}$ following the assumptions of Theorem 1, $r_t^K \leqslant r_t^{(0)} \leqslant 2\sqrt{\beta_t}\sigma_t^0 + \frac{1}{t^2}$, holds with probability at least $1 - \delta$.*

We refer to [31] (Lemmas 5.2, 5.8) for the detailed proof of the bound for $r_t^{(0)}$.

Now we show an intermediate result bounding the deviations at the points x_{t+1}^0 by the one at the points x_t^{K-1}.

Lemma 2. *The deviation of the point selected by the UCB policy is bounded by the one for the last point selected by the PE policy at the previous iteration, whp, $\forall t < T, \ \sigma_{t+1}^0 \leqslant \sigma_t^{K-1}$*

Proof. By the definitions of x_{t+1}^0 (Eq.7), we have $\widehat{f}_{t+1}^+(x_{t+1}^0) \geqslant \widehat{f}_{t+1}^+(x_t^\bullet)$. Then, we know with high probability that $\forall x \in \mathcal{X}, \forall t < T, \ \widehat{f}_{t+1}^+(x) \geqslant \widehat{f}_t^-(x)$. We can therefore claim *whp* $\widehat{f}_{t+1}^+(x_{t+1}^0) \geqslant y_t^\bullet$, and thus that $x_{t+1}^0 \in \mathfrak{R}_t^+$ *whp*.

We have as a result by the definition of x_t^{k-1} (Eq.9) that $\widehat{\sigma}_t^{(k-1)}(x_{t+1}^0) \leqslant \widehat{\sigma}_t^{(k-1)}(x_t^{k-1})$ *whp*. Using the "Information never hurts" principle [25], we know that the entropy of $f(x)$ for all location x decreases while we observe points x_t. For GP, the entropy is also a non-decreasing function of the variance, so that

$$\forall x \in \mathcal{X}, \ \widehat{\sigma}_{t+1}^{(0)}(x) \leqslant \widehat{\sigma}_t^{(k-1)}(x) \ .$$

We thus prove $\sigma_{t+1}^0 \leqslant \sigma_t^{k-1}$.

Lemma 3. *The sum of the deviations of the points selected by the UCB policy are bounded by the one for all the selected points divided by K, whp,*

$$\sum_{t=0}^{T-1} \sigma_t^0 \leqslant \frac{1}{K} \sum_{t=0}^{T-1} \sum_{k=0}^{K-1} \sigma_t^k \ .$$

Proof. Using Lemma 2 and the definitions of the x_t^k, we have that $\sigma_{t+1}^0 \leqslant \sigma_t^k$ for all $k \geqslant 1$. Summing over k, we get for all $t \geqslant 0$, $\sigma_t^0 + (K-1)\sigma_{t+1}^0 \leqslant \sum_{k=0}^{K-1} \sigma_t^k$. Now, summing over t and with $\sigma_0^0 \geqslant 0$ and $\sigma_T^0 \geqslant 0$, we obtain the desired result.

Next, we can bound the sum of all posterior variances $(\sigma_t^k)^2$ via the maximum information gain for a sequence of TK locations.

Lemma 4. *The sum of the variances of the selected points are bounded by a constant factor times γ_{TK}, $\exists C_1' \in \mathbb{R}$, $\sum_{t<T} \sum_{k<K} (\sigma_t^k)^2 \leqslant C_1' \gamma_{TK}$ where γ_{TK} is the maximum information gain obtainable by a sequential procedure of length TK.*

Proof. We know that the information gain for a sequence of T locations x_t can be expressed in terms of the posterior variances $(\widehat{\sigma}_{t-1}(x_t))^2$. The deviations σ_t^k being independent of the observations y_t^k, the same equality holds for the posterior variances $(\widehat{\sigma}_t^{(k)}(x_t^k))^2$. See Lemmas 5.3 and 5.4 in [31] for the detailed proof, giving $C_1' = \frac{2}{\log(1+\sigma^{-2})}$.

Lemma 5. *The cumulative regret can be bound in terms of the maximum information gain, whp, $\exists C_1, C_2 \in \mathbb{R}$,*

$$\sum_{t<T} r_t^K \leqslant \sqrt{\frac{T}{K} C_1 \beta_T \gamma_{TK} + C_2} \ .$$

Proof. Using the previous lemmas and the fact that $\beta_t \leqslant \beta_T$ for all $t \leqslant T$, we have in the case of finite $\mathcal{X}$, *whp*

$$\sum_{t<T} r_t^K \leqslant \sum_{t<T} 2\sqrt{\beta_t}\sigma_t^0 \text{ , by Lemma 1}$$

$$\leqslant 2\sqrt{\beta_T}\frac{1}{K}\sum_{t<T}\sum_{k<K}\sigma_t^k \text{ , by Lemma 3}$$

$$\leqslant 2\sqrt{\beta_T}\frac{1}{K}\sqrt{TK\sum_{t<T}\sum_{k<K}(\sigma_t^k)^2} \text{ , by Cauchy-Schwarz}$$

$$\leqslant 2\sqrt{\beta_T}\frac{1}{K}\sqrt{TKC_1'\gamma_{TK}} \text{ , by Lemma 4}$$

$$\leqslant \sqrt{\frac{T}{K}C_1\beta_T\gamma_{TK}} \text{ with } C_1 = \frac{4}{\log(1+\sigma^{-2})} \ .$$

For compact and convex $\mathcal{X}$, a similar reasoning gives

$$R_T^K \leqslant \sqrt{\frac{T}{K}C_1\beta_T\gamma_{TK} + C_2} \text{ with } C_2 = \frac{\pi}{\sqrt{6}} < 2 \ .$$

Lemma 5 conclude the proof of Theorem 1 for the regret R_T^K. The analysis for R_{TK} is simpler, using the Lemma 6 which bounds the regret for the Pure Exploration queries, leading to $C_1 = \frac{36}{\log(1+\sigma^{-2})}$.

Lemma 6. *The regret for the queries x_t^k selected by Pure Exploration in $\mathfrak{R}_t^+$ are bounded whp by, $6\sqrt{\beta_t}\sigma_t^k$.*

Proof. As in Lemma 1, we have *whp*, for all $t \leqslant T$ and $k \geqslant 1$,

$$r_t^{(k)} \leqslant \widehat{\mu}_t(x^\star) + \sqrt{\beta_t}\widehat{\sigma}_t(x^\star) - \widehat{\mu}(x_t^k) + \sqrt{\beta_t}\sigma_t^k$$

$$\leqslant f(x_t^\bullet) + 2\sqrt{\beta_t}\widehat{\sigma}_t(x^\star) - \widehat{\mu}(x_t^k) + \sqrt{\beta_t}\sigma_t^k \text{ by definition of } x_t^\bullet$$

$$\leqslant \widehat{\mu}(x_t^k) + 2\sqrt{\beta_{t+1}}\sigma_t^k + 2\sqrt{\beta_t}\sigma_t^k - \widehat{\mu}(x_t^k) + \sqrt{\beta_t}\sigma_t^k \text{ by definition of } \mathfrak{R}_t^+$$

$$\leqslant 3\sqrt{\beta_t}\sigma_t^k + 2\sqrt{\beta_t}\sigma_t^k + \sqrt{\beta_t}\sigma_t^k \text{ by definition of } \beta_{t+1}$$

$$\leqslant 6\sqrt{\beta_t}\sigma_t^k$$

To conclude the analysis of R_{TK} and prove Theorem 1, it suffices to use then the last three steps of Lemma 5.

5   Experiments

5.1   Protocol

We compare the empirical performances of our algorithm against the state of the art of global optimization by batches, GP-BUCB [13] and SM-UCB [4]. The tasks used for assessment come from three real applications and two synthetic
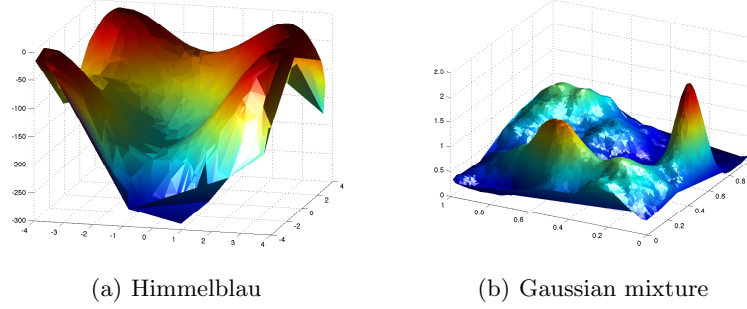
(a) Himmelblau   (b) Gaussian mixture

Fig. 3. Visualization of the synthetic functions used for assessment

problems described here. The results are shown in Figure 4. For all datasets and algorithms, the size of the batches K was set to 10 and the learners were initialized with a random subset of 20 observations (x_i, y_i). The curves on Figure 4 show the evolution of the regret R_t^K in term of iteration t. We report the average value with the confidence interval over 64 experiments. The parameters of the algorithm, like the bandwidth of the RBF Kernel were chosen as the best parameters found by maximization of the posterior likelihood.

5.2   Description of data sets

Generated GP. The `Generated GP` functions are random GPs drawn from a Matérn kernel (Eq. 3) in dimension 2, with the kernel bandwidth set to $\frac{1}{4}$, the Matérn parameter $\nu = 3$ and noise variance σ^2 set to 1.

Gaussian Mixture. This synthetic function comes from the addition of three 2-D Gaussian functions. at $(0.2, 0.5)$, $(0.9, 0.9)$, and the maximum at $(0.6, 0.1)$. We then perturb these Gaussian functions with smooth variations generated from a Gaussian Process with Matérn Kernel and very few noise. It is shown on Figure 3(b). The highest peak being thin, the sequential search for the maximum of this function is quite challenging.

Himmelblau function. The `Himmelblau` task is another synthetic function in dimension 2. We compute a slightly tilted version of the Himmelblau's function, and take the opposite to match the challenge of finding its maximum. This function presents four peaks but only one global maximum. It gives a practical way to test the ability of a strategy to manage exploration/exploitation tradeoffs. It is represented in Figure 3(a).

Mackey-Glass function. The Mackey-Glass delay-differential equation [2] is a chaotic system in dimension 6, but without noise. It models real feedback systems

---

[2] `http://www.scholarpedia.org/article/Mackey-Glass_equation`

(a) Generated GP   (b) Himmelblau   (c) Gaussian Mixture

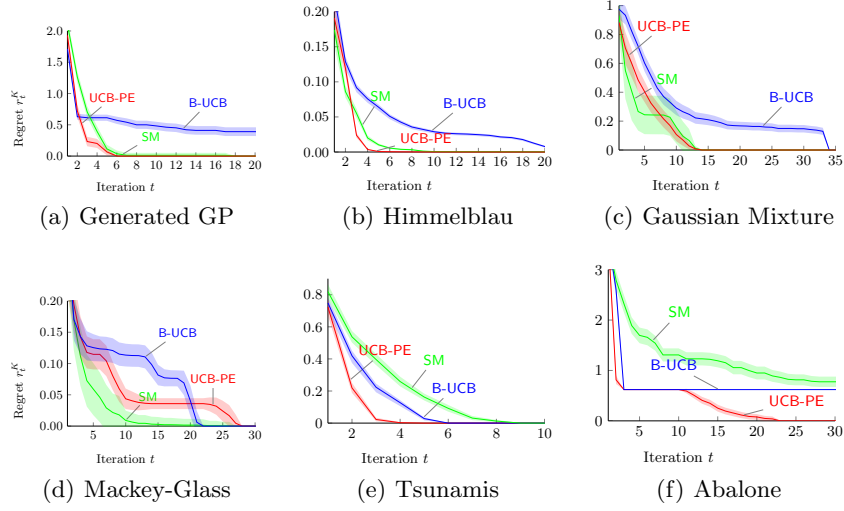(d) Mackey-Glass   (e) Tsunamis   (f) Abalone

Fig. 4. Experiments on several real and synthetics tasks. The curves show the decay of the mean of the simple regret r_t^K with respect to the iteration t, over 64 experiments. We show with the translucent area the confidence intervals.

and is used in physiological domains such as hematology, cardiology, neurology, and psychiatry. The highly chaotic behavior of this function makes it an exceptionally difficult optimization problem. It has been used as a benchmark for example by [16].

Tsunamis. Recent post-tsunami survey data as well as the numerical simulations of [21] have shown that in some cases the run-up, which is the maximum vertical extent of wave climbing on a beach, in areas which were supposed to be protected by small islands in the vicinity of coast, was significantly higher than in neighboring locations. Motivated by these observations [32] investigated this phenomenon by employing numerical simulations using the VOLNA code [14] with the simplified geometry of a conical island sitting on a flat surface in front of a sloping beach. Their setup was controlled by five physical parameters and their aim was to find with confidence and with the least number of simulations the maximum run-up amplification on the beach directly behind the island, compared with the run-up on a lateral location, not influenced by the presence of the island. Since this problem is too complex to treat analytically, the authors had to solve numerically the Nonlinear Shallow Water Equations.

Abalone. The challenge of the `Abalone` dataset is to predict the age of a specie of sea snails from physical measurements. It comes from the study by [29] and it is provided by the UCI Machine Learning Repository [3]. We use it as a maximization problem in dimension 8.

---

[3] `http://archive.ics.uci.edu/ml/datasets/Abalone`

5.3   Comparison of algorithms

The algorithm SM —Simulation Matching— described in [4], with UCB base policy, has shown similar results to GP-UCB-PE on synthetic functions (Figures 4(a), 4(b), 4(c)) and even better results on chaotic problem without noise (Figure 4(d)), but performs worse on real noisy data (Figures 4(e), 4(f)). On the contrary, the initialization phase of GP-BUCB leads to good regret on difficult real tasks (Figure 4(e)), but looses time on synthetic Gaussian or polynomial ones (Figures 4(a), 4(b), 4(c)). The number of dimensions of the `Abalone` task is already a limitation for GP-BUCB with the RBF kernel, making the initialization phase time-consuming. The mean regret for GP-BUCB converges to zero abruptly after the initialization phase at iteration 55, and is therefore not visible on Figure 4(f), as for 4(c) where its regret decays at iteration 34.

GP-UCB-PE achieves good performances on both sides. We obtained better regret on synthetic data as well as on real problems from the domains of physics and biology. Moreover, the computation time of SM was two order of magnitude longer than the others.

6   Conclusion

We have presented the GP-UCB-PE algorithm which addresses the problem of finding in few iterations the maximum of an unknown arbitrary function observed via batches of K noisy evaluations. We have provided theoretical bounds for the cumulative regret obtained by GP-UCB-PE in the Gaussian Process settings. Through parallelization, these bounds improve the ones for the state-of-the-art of sequential GP optimization by a ratio of $\sqrt{K}$, and are strictly better than the ones for GP-BUCB, a concurrent algorithm for parallel GP optimization. We have compared experimentally our method to GP-BUCB and SM-UCB, another approach for parallel GP optimization lacking of theoretical guarantees. These empirical results have confirmed the effectiveness of GP-UCB-PE on several applications.

The strategy of combining in the same batch some queries selected via Pure Exploration is an intuitive idea that can be applied in many other methods. We expect for example to obtain similar results with the Maximum Expected Improvement policy (MEI). Any proof of regret bound that relies on the fact that the uncertainty decreases with the exploration should be easily adapted to a paralleled extension with Pure Exploration.

On the other hand, we have observed in practice that the strategies which focus more on exploitation often lead to faster decrease of the regret, for example the strategy that uses K times the GP-UCB criterion with updated variance. We conjecture that the regret for this strategy is unbounded for general GPs, justifying the need for the initialization phase of GP-BUCB. However, it would be relevant to specify formally the assumptions needed by this greedy strategy to guarantee good performances.

Bibliography

[1] Audibert, J.Y., Bubeck, S., Munos, R.: Bandit view on noisy optimization. In: Optimization for Machine Learning, pp. 431–454. MIT Press (2011)

[2] Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multi-armed bandit problem. Machine Learning 47(2-3), 235–256 (2002)

[3] Auer, P., Ortner, R., Szepesvári, C.: Improved rates for the stochastic continuum-armed bandit problem. In: Proceedings of COLT. pp. 454–468. Omnipress (2007)

[4] Azimi, J., Fern, A., Fern, X.: Batch bayesian optimization via simulation matching. In: Advances in NIPS. pp. 109–117. Curran Associates, Inc. (2010)

[5] Bubeck, S., Cesa-Bianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends in Machine Learning 5(1), 1–122 (2012)

[6] Bubeck, S., Munos, Stoltz, G., Szepesvári, C.: Online optimization in x-armed bandits. In: Advances in NIPS. pp. 201–208. Curran Associates, Inc. (2008)

[7] Bubeck, S., Munos, R., Stoltz, G.: Pure exploration in multi-armed bandits problems. In: Proceedings of ALT, pp. 23–37. Springer-Verlag (2009)

[8] Carpentier, A., Lazaric, A., Ghavamzadeh, M., Munos, R., Auer, P.: Upper-confidence-bound algorithms for active learning in multi-armed bandits. In: Proceedings of the ALT. pp. 189–203. Springer-Verlag (2011)

[9] Chen, B., Castro, R., Krause, A.: Joint optimization and variable selection of high-dimensional gaussian processes. In: Proceedings of ICML. ACM (2012)

[10] Chen, Y., Krause, A.: Near-optimal batch mode active learning and adaptive submodular optimization. In: Proceedings of ICML. ACM (2013)

[11] Coquelin, P.A., Munos, R.: Bandit algorithms for tree search. In: Proceedings of UAI. pp. 67–74. AUAI Press (2007)

[12] Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley-Interscience (1991)

[13] Desautels, T., Krause, A., Burdick, J.: Parallelizing exploration-exploitation tradeoffs with gaussian process bandit optimization. In: Proceedings of ICML. icml.cc / Omnipress (2012)

[14] Dutykh, D., Poncet, R., Dias, F.: The VOLNA code for the numerical modelling of tsunami waves: generation, propagation and inundation. European Journal of Mechanics B/Fluids 30, 598–615 (2011)

[15] Fedorov, V.V.: Theory of Optimal Experiments. Academic Press (1972)

[16] Flake, G.W., Lawrence, S.: Efficient svm regression training with smo. Machine Learning 46(1-3), 271–290 (2002)

[17] de Freitas, N., Smola, A.J., Zoghi, M.: Exponential regret bounds for gaussian process bandits with deterministic observations. In: Proceedings of ICML. icml.cc / Omnipress (2012)

[18] Grünewälder, S., Audibert, J.Y., Opper, M., Shawe-Taylor, J.: Regret Bounds for Gaussian Process Bandit Problems. In: Proceedings of AIS-TATS. pp. 273–280. MIT Press (2010)

[19] Guestrin, C., Krause, A., Singh, A.: Near-optimal sensor placements in Gaussian processes. In: Proceedings of ICML. pp. 265–272. ACM (2005)

[20] Hennig, P., Schuler, C.J.: Entropy search for information-efficient global optimization. Journal of Machine Learning Research 13, 1809–1837 (2012)

[21] Hill, E.M., Borrero, J.C., Huang, Z., Qiu, Q., Banerjee, P., Natawidjaja, D.H., Elosegui, P., Fritz, H.M., Suwargadi, B.W., Pranantyo, I.R., Li, L., Macpherson, K.A., Skanavis, V., Synolakis, C.E., Sieh, K.: The 2010 mw 7.8 mentawai earthquake: Very shallow source of a rare tsunami earthquake determined from tsunami field survey and near-field gps data. J. Geophys. Res. 117, B06402– (2012)

[22] Kleinberg, R.: Nearly tight bounds for the continuum-armed bandit problem. In: Advances in NIPS. pp. 697–704. MIT Press (2004)

[23] Ko, C., Lee, J., Queyranne, M.: An exact algorithm for maximum entropy sampling. Operations Research pp. 684–691 (1995)

[24] Kocsis, L., Szepesvári, C.: Bandit based monte-carlo planning. In: Proceedings of ECML. pp. 282–293. Springer (2006)

[25] Krause, A., Guestrin, C.: Near-optimal nonmyopic value of information in graphical models. In: Proceedings of UAI. pp. 324–331. AUAI Press (2005)

[26] Kuss, M., Pfingsten, T., Csat, L., Rasmussen, C.E., Kuss, M., Pfingsten, T., Csat, L., Rasmussen, C.E.: Approximate inference for robust gaussian process regression (2005)

[27] Mes, M.R., Powell, W.B., Frazier, P.I.: Hierarchical knowledge gradient for sequential sampling. Journal of Machine Learning Research 12, 2931–2974 (2011)

[28] Mockus, J.: Bayesian approach to global optimization: theory and applications. Mathematics and its applications (Kluwer Academic Publishers).: Soviet series, Kluwer Academic (1989)

[29] Nash, W., Tasmania. Marine Research Laboratories: The population biology of abalone (haliotis species) in tasmania: Blacklip abalone (h. rubra) from the north coast and the islands of bass strait. Tech. rep., Tasmania. Sea Fisheries Division (1994)

[30] Rasmussen, C.E., Williams, C.: Gaussian Processes for Machine Learning. MIT Press (2005)

[31] Srinivas, N., Krause, A., Kakade, S., Seeger, M.: Gaussian process optimization in the bandit setting: No regret and experimental design. In: Proceedings of ICML. pp. 1015–1022. Omnipress (2010)

[32] Stefanakis, T.S., Dias, F., Vayatis, N., Guillas, S.: Long-wave runup on a plane beach behind a conical island. In: Proceedings of WCEE (2012)

[33] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. A Bradford Book (1998)